

J. R. Statist. Soc. A (2019)
182, Part 2, pp. 403–418

Graphics for uncertainty

Adrian W. Bowman

University of Glasgow, UK

[Read before The Royal Statistical Society at a meeting on 'Data visualization' at the Society's 2018 annual conference in Cardiff on Wednesday, September 5th, 2018, the President, Professor D. J. Spiegelhalter, in the Chair]

Summary. Graphical methods such as colour shading and animation, which are widely available, can be very effective in communicating uncertainty. In particular, the idea of a 'density strip' provides a conceptually simple representation of a distribution and this is explored in a variety of settings, including a comparison of means, regression and models for contingency tables. Animation is also a very useful device for exploring uncertainty and this is explored particularly in the context of flexible models, expressed in curves and surfaces whose structure is of particular interest. Animation can further provide a helpful mechanism for exploring data in several dimensions. This is explored in the simple but very important setting of spatiotemporal data.

Keywords: Animation; Density strip; Graphics; Spatiotemporal data; Uncertainty

1. Introduction

In 2010, the Royal Statistical Society launched a 10-year statistical literacy campaign with a discussion paper by Wild *et al.* (2011) on making statistical concepts accessible. It is stating the obvious to say that graphical methods play a very important role in both the communication of statistical information and concepts in a manner which is largely free of technical language. There is, of course, a long tradition of the development of innovation in statistical graphics. Examples include exploratory data analysis (Tukey, 1977), the careful visual design expressed in lattice graphics (Sarkar, 2008) and the animation and interaction provided by systems such as *XLisp-Stat* (Tierney, 1988), *ggobi* (Cook and Swayne, 2007) and *Mondriaan* (Theus and Urbanek, 2008). The *ggplot2* system (Wickham, 2009), based on the 'grammar of graphics' (Wilkinson, 2005), is proving increasingly popular through its combination of attractive graphical design and flexibility. The wide variety of tools that are now available to general users is illustrated in Chen *et al.* (2008). There are also general tools for enabling user interaction, provided for example in systems such as *Shiny* (Chang *et al.*, 2016).

Despite this array of tools, the standard approaches to graphical display remain those based on relatively simple point and line drawings, such as histograms, boxplots, bar charts and scatter plots, supplemented occasionally by colour filling. However, to those who are unfamiliar with statistical methods, the carefully positioned lines of a boxplot, or the boundaries of a confidence interval, imply a precision and encourage an algorithmic approach to the evaluation of evidence, which is at odds with the concept of uncertainty and variation. Of course, technicality and precision are very important but the communication of results, and more informal evaluation

Address for correspondence: Adrian W. Bowman, School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8QQ, UK.
E-mail: adrian.bowman@glasgow.ac.uk

© 2019 The Author Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/19/182403
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

of the statistical evidence that they encapsulate, is greatly aided by graphical methods which clearly communicate the uncertainties that are involved.

The focus of this paper is on graphical effects which can add considerable value in the exploration of data and models in general and in the quantification of uncertainty in particular. An example is provided by Jackson (2008) who introduced the concept of a *density strip* to represent distributions by the simple device of a bar of colour whose intensity is proportional to the density function at each location. This produces a visually appealing display whose continuous gradations match well with the intuition of what uncertainty means. This is illustrated in Fig. 1(a) which displays an animation. (Animated displays of this type are used repeatedly through the paper.) Observations are displayed successively, in a manner which represents the unpredictable '*there it is—no, there it is*' nature of random variation. Above the display of individual data points, the accumulated 'footprints' of the observations build into a density strip, where intensity reflects the relative frequency of observations at different locations. Use of a footprint with faded edges, centred on the location of each observation, produces a smooth kernel density estimate in a manner which can be appreciated entirely intuitively (with the role of the smoothing parameter refreshingly expressed in shoe size!). Fig. 1(b) represents several distributions of standard form, where features such as skewness and bimodality are all expressed in a perfectly clear but very compact manner. Fig. 1(c) displays density estimates of data on the characteristics of aircraft design from 1914 to 1984, described by Bowman and Azzalini (1997), using $\log(\text{Speed})$ separated into three time periods (1, 1914–1935; 2, 1936–1955; 3, 1956–1984). The shifting locations and sizes of the components are apparent. The ability to draw a distribution along a single axis is particularly appealing. This fits well with the recommendation of Wild *et al.* (2011) that plots should 'stay in the same visual space' as the data, in contrast with a histogram which requires a second axis to display frequency.

Another graphical device which can be used to good effect is animation—a device that has already been used for Fig. 1. If animation is not carefully employed it may simply be a distraction. Tversky *et al.* (2002) reviewed its use from a perceptual perspective and identified that it was likely to be most effective in displaying changes over space and time, particularly when user interaction was available. This is very often precisely the context in which animation is used in statistical graphics. Hullman *et al.* (2015) carried out a detailed study of displays of probability distributions employing static and animated graphics and concluded that the latter assisted in the perception of distributional properties. It is clear from the wide differences in individual perceptions across viewers that no graphic display will work effectively for everyone, but it is also clear that animation is a form of display which is effective for many.

This paper explores how these relatively simple graphical devices of colour shading and animation can assist in the display of data, models and the associated uncertainty. The methods are particularly geared towards communicating with those who do not have technical statistical knowledge, but the aim is to do so in ways which are directly compatible with more technical forms of analysis. Section 2 exploits density strips in a simple comparison of means and extends this to the assessment of evidence for interaction in analysis of variance and of association in contingency tables. Section 3 deals with standard regression models whereas Section 4 considers how the uncertainty that is associated with the estimation of flexible curves, and particularly surfaces, can be displayed through animation. Section 5 uses both colour shading and animation in addressing the issue of displaying spatiotemporal data, which are now a commonly occurring data structure, and associated spatiotemporal models. Some final discussion is given in Section 6.

The data for the examples and code to construct the graphical displays can be obtained from <http://dx.doi.org/10.5525/gla.researchdata.595>.

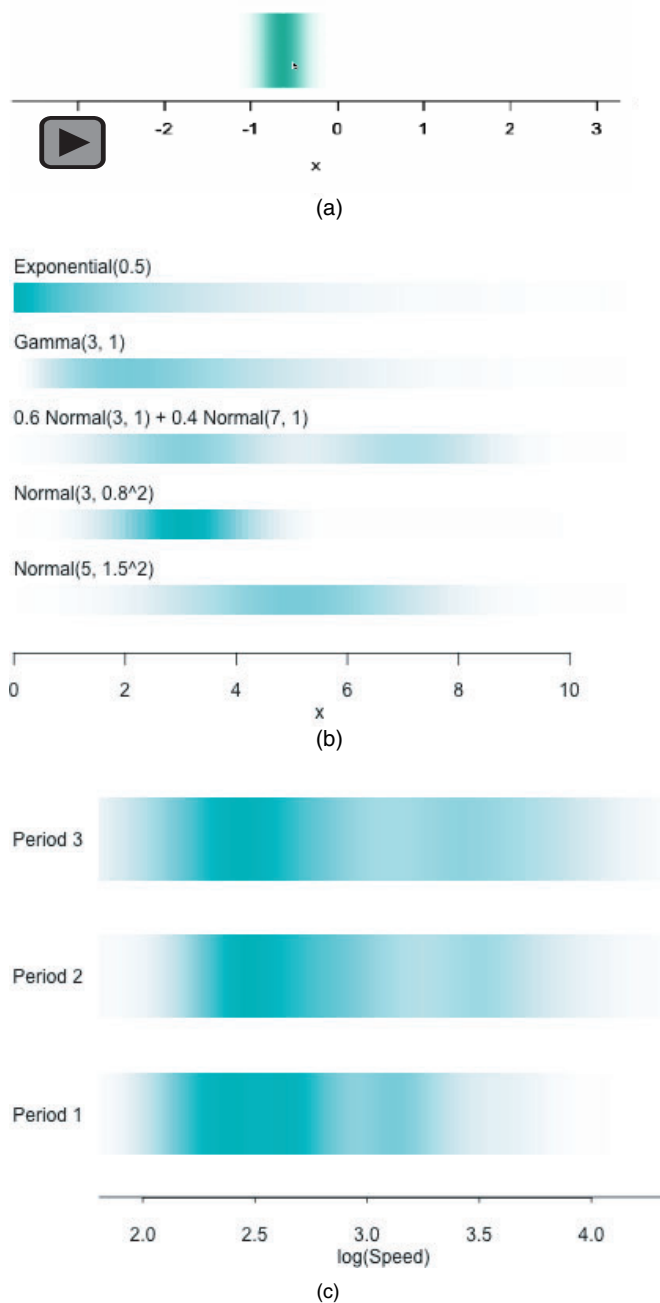


Fig. 1. Animation (a) illustrates the unpredictable nature of sampled observations and the accumulation of the 'footprints' of the data into a density strip representation of a density estimate whereas (b) shows several distributions in density strip form and (c) shows density strips constructed from data on aircraft speed, on a log-scale, for three different time periods

2. Comparing models

2.1. Comparing means

Fig. 2(a) displays data on changes in asymmetry scores derived from the facial images of children in a study of the effects of corrective surgery on patients who were born with a unilateral cleft lip or unilateral cleft lip and palate. The context of the study is described by Hood *et al.* (2004) and the methodology for measuring asymmetry is described by Bock and Bowman (2006). The data that are plotted here show the change in asymmetry scores from facial images captured at 3 months and 6 months of age. A substantial change in asymmetry is expected for cleft cases over this period, as this is when corrective surgery takes place. Interest here lies in whether there is any change in the mean asymmetry scores for controls. The differences (3 months – 6 months) are plotted, with a grey density strip to highlight the underlying distribution and with the sample mean \bar{x} and reference value 0 marked as lines. The data are shown as points, with added vertical ‘jitter’ simply to avoid overplotting.

The middle strip in Fig. 2(a) highlights the uncertainty in estimation of the true mean μ . In a Bayesian analysis, the posterior distribution of the parameter of interest encapsulates all the relevant information and a density strip provides a simple graphical expression of this. From a frequentist perspective, with se denoting standard error and n denoting sample size, the key step in the argument is the distribution of the pivotal quantity $(\bar{x} - \mu)/se(\bar{x})$, which has a t_{n-1} -distribution when an assumption of normality is appropriate. Instead of proceeding to a confidence interval through the usual inversion argument, scaling the t -distribution (or its standard normal approximation) by the standard error expresses the distribution of the distance between \bar{x} and μ . This may be regarded as a kind of ‘ruler’ which measures variation and which, when placed at \bar{x} , provides a graphical expression of uncertainty. This has a direct correspondence with equivalent confidence intervals but it avoids the rather sophisticated interpretation that is required for a detailed derivation. The middle strip of Fig. 2(a) provides a clear indication that, when uncertainty is taken into account, the evidence of a change in mean asymmetry score from 3 months to 6 months is not convincing. An alternative approach is to centre the distribution at the reference value of 0, as in the lower strip of Fig. 2(a), to express uncertainty in the position of \bar{x} under the assumption that the true mean is 0. This provides a simple graphical expression of the essential concepts of a hypothesis test, without the need for complex explanations of p -values. (For the record, the p -value here is 0.17.) This point was well made by Jackson (2008), who provided numerous additional examples of the helpful uses of density strips.

Fig. 2(b) extends this to the two-sample setting, here in the comparison of asymmetry scores for unilateral cleft lip and unilateral cleft lip and palate patients at 3 months. The higher mean score for the unilateral cleft lip and palate group is apparent (for the record, $p < 0.001$), with the central density strip expressing the uncertainty in the size of the difference in mean scores. By locating the 0-value on the difference scale at the smaller of the sample means, the graphical display gives a clear indication of the plausible size of the difference. The linking of the two scales aids interpretation and respects the principle that was advocated by Wild *et al.* (2011) that graphics should remain in the same ‘visual space’ as the data.

The intention of these plots is to convey uncertainty through graphics which are based on the usual technical constructs but which allow inferences to be drawn, and the size and nature of effects identified, in an informal and conceptual manner. In particular, the ‘fuzzy’ nature of a density strip aligns with an intuitive concept of uncertainty more naturally than the precise end points of confidence intervals.

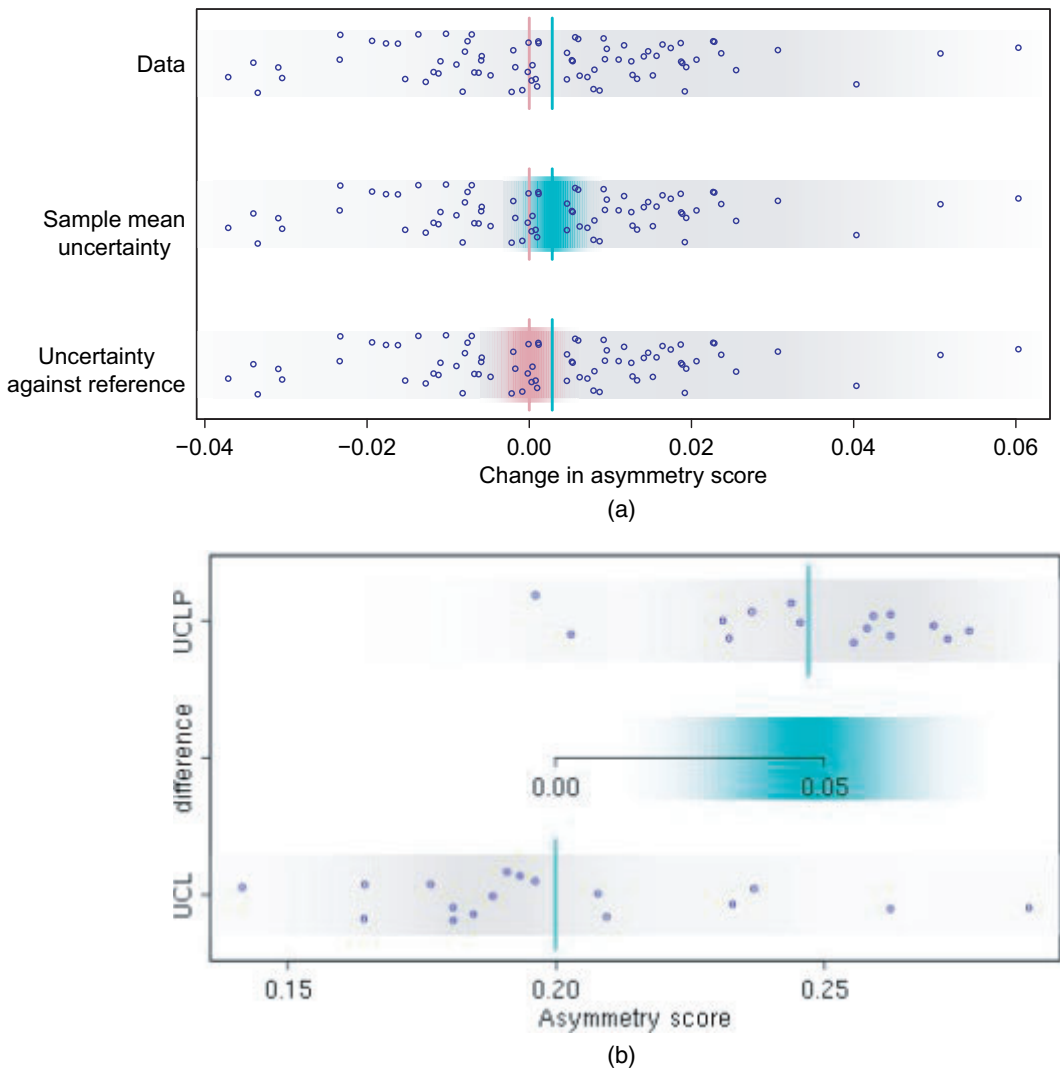


Fig. 2. (a) Data (○) on the change in asymmetry score of control children from 3 months to 6 months (■, distribution of the data; |, sample mean; |, reference point of 0, corresponding to no change) (the middle strip indicates the uncertainty associated with the sample mean whereas the lower strip locates this at the reference value of 0) and (b) asymmetry scores (○) at 3 months from unilateral cleft lip and unilateral cleft lip and palate patient groups (|, sample means; ■, measure of uncertainty of the difference in means)

2.2. Factor models

Fig. 3 displays data on sulphur dioxide (SO_2) air pollution measured on a log-scale at three European sites in three different years. Bowman *et al.* (2009) described the wider data set. Fig. 3(a) displays grey density strips for each data group, with red lines superimposed to indicate the fitted values from a linear model with site, year and interaction effects.

A simple two-way analysis of variance allows evidence for the presence of interaction to be explored by the construction of an F -statistic which contrasts the residual sums of squares from an additive and an interaction model. In this form, the F -statistic is rather remote from the graphical representation of the data in Fig. 3 but the algebra of linear models (Seber, 1977)

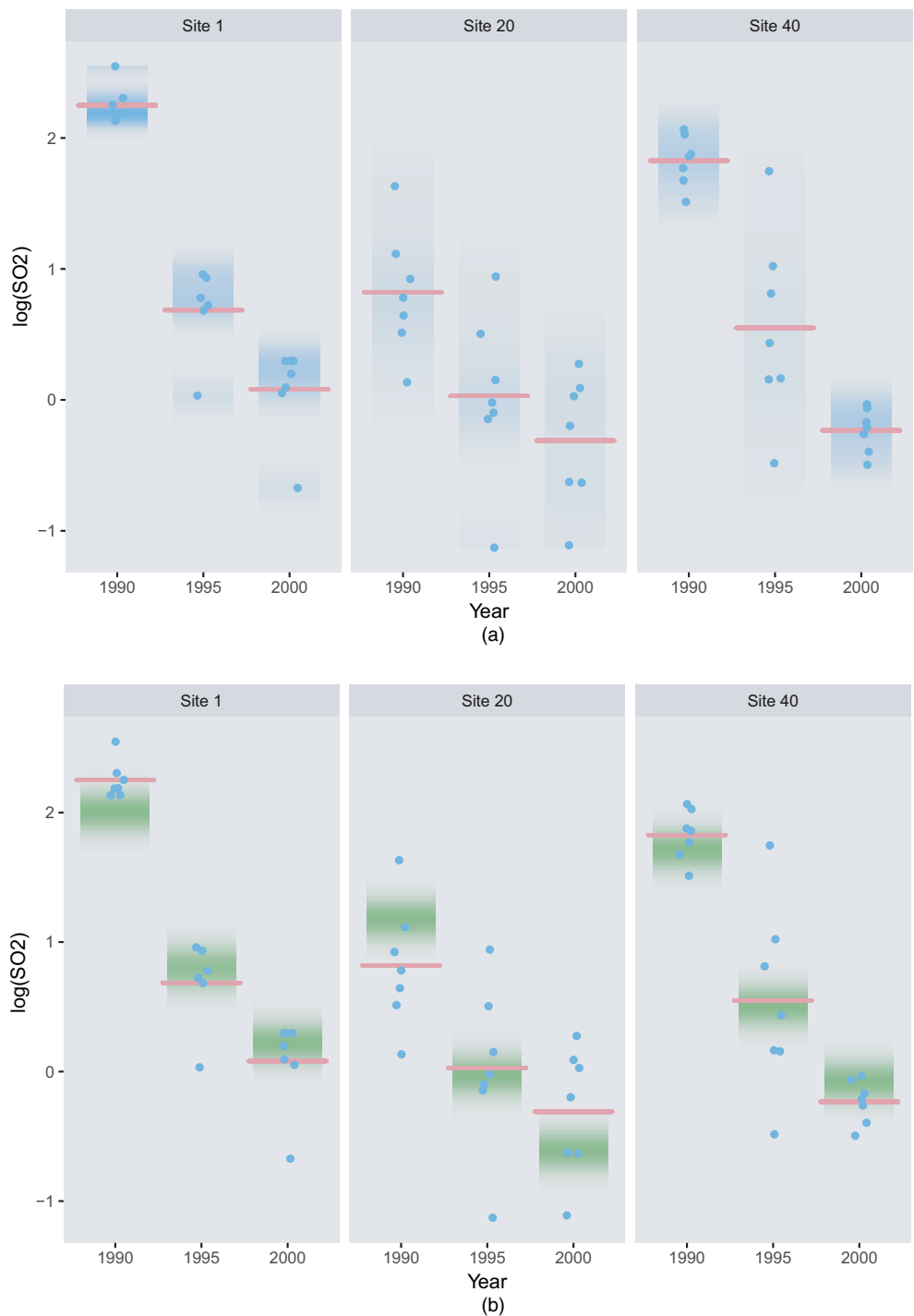


Fig. 3. (a) Data on SO_2 pollution, on a log-scale with density strips to highlight the groups (—, fitted values from an interaction model), and (b) density strips to express the uncertainty in the comparison with the simpler additive model

enables this to be expressed as a comparison of the fitted values from the two models. For fitted values $\{\hat{y}_{ij}, \hat{y}_{ij0} : i = 1, \dots, I; j = 1, \dots, J\}$, where the subscript 0 indicates the simpler additive model, the F -statistic becomes

$$F = \frac{\sum_{i,j} n_{ij} (\hat{y}_{ij} - \hat{y}_{ij0})^2 / \nu}{\hat{\sigma}^2} = \frac{1}{IJ} \sum_{i,j} \left[\frac{\hat{y}_{ij} - \hat{y}_{ij0}}{\sqrt{\{\nu/(IJ)\} \hat{\sigma}^2 / n_{ij}}} \right]^2 \quad (1)$$

where ν denotes the difference in the degrees of freedom for the two models and $\hat{\sigma}$ denotes the estimate of error standard deviation from the larger model.

This is expressed graphically in Fig. 3(b), where normal distributions, centred on the fitted values of the additive model and with standard deviations

$$\frac{\sqrt{\{\nu/(IJ)\} \hat{\sigma}^2}}{\sqrt{n_{ij}}},$$

are represented as density strips. These characterize the location, with uncertainty, of the fitted values of the additive model, against which the fitted values of the interaction model can be compared. The focus here is not on representing uncertainty through the standard errors of the comparison of fitted values at each factor combination, with suitable adjustment for the multiple comparisons that are involved, but instead to indicate the individual contributions to the overall assessment of evidence for the suitability of an additive model through the F -statistic. The marked mismatch, on average, between the fitted values from the interaction model and the corresponding ranges of values consistent with the additive model gives graphical expression to the evidence that interaction is present. This can, of course, be made more precise by comparing the observed value of the F -statistic (3.18) with the $F_{4,53}$ -distribution. (There are 62 observations in the data set.) To strengthen further the link with the graphical comparison of fitted values, the test statistic and its reference distribution can be expressed on a square-root scale, with an interpretation as the root-mean-square differences of the fitted values from the two models. The density strip representation of this,



indicates a strong degree of mismatch, which is entirely consistent with the technical details of the underlying F -test (which, for the record, produces a p -value of 0.020 in this example).

The aim of these examples is not to promote the use of significance tests, whose overuse has rightly been criticized and whose interpretation is often misunderstood. Instead, the aim is to advocate the use of a reference model against which the size and nature of effects of interest can be evaluated, using appropriate graphics to provide measures of uncertainty. These graphics are consistent with standard methods of analysis of variance. The comparison of fitted values expressed in equation (1) can clearly be extended to a wider range of linear models.

2.3. Contingency tables

Graphical methods for data in the form of contingency tables include ‘mosaic plots’ (Hartigan and Kleiner, 1981). A wider variety of approaches are discussed in Friendly (2000) and Gelman *et al.* (2002). A famous example of a contingency table is provided by the landmark study on the association between smoking and lung cancer, which was conducted by Doll and Hill (1950), where the prevalence of smoking was examined both in lung cancer patients and in a group of controls consisting of patients suffering from diseases other than cancer. The vast

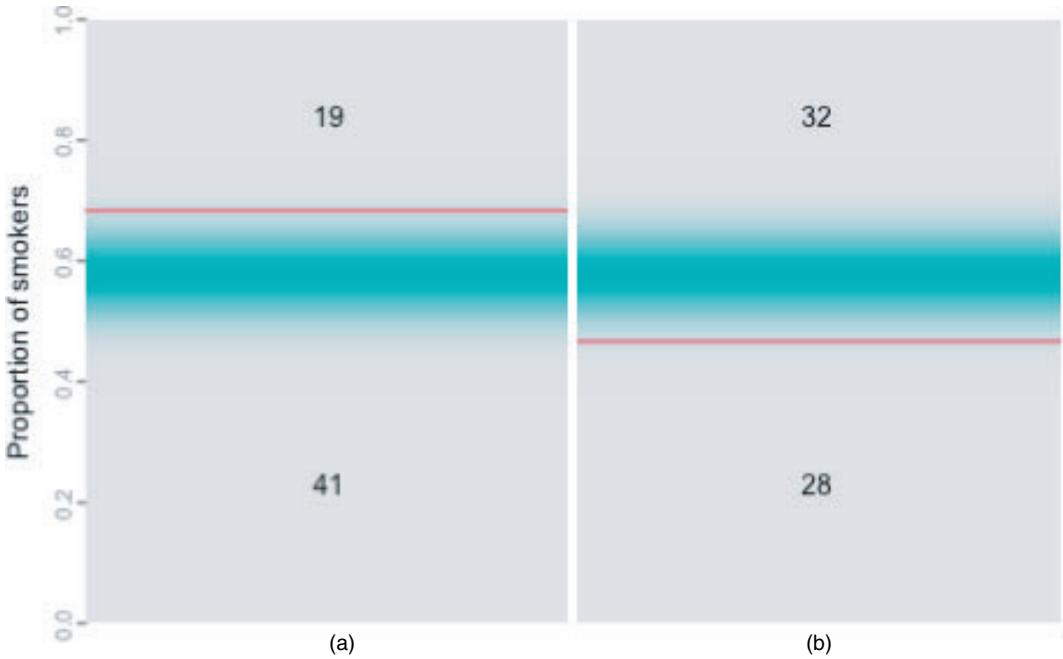


Fig. 4. Graphical display of data on (a) cases and (b) controls from a study of smoking and lung cancer reported by Doll and Hill (1950), with uncertainty enabling the presence of association to be assessed

majority of patients were men but Fig. 4 shows the data for women. The columns correspond to fixed sample sizes (60) drawn from the case and control populations, whereas the horizontal lines mark the observed proportions of smokers in each group. This is a simple form of mosaic plot where observed counts are represented by the areas of the four displayed regions whereas the vertical axis focuses attention on the proportion scale, which is most relevant for comparison.

If the counts over rows $\{i: 1, \dots, I\}$ and columns $\{j: 1, \dots, J\}$ are denoted by n_{ij} and the dot notation denotes summation over the subscript indicated, then the χ^2 -statistic for a test of no association can be rewritten as

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - n_{i.}n_{.j}/n_{..})^2}{n_{i.}n_{.j}/n_{..}} = \frac{1}{IJ} \sum_{ij} \left[\frac{\hat{p}_{ij} - \hat{p}_{i0}}{\sqrt{\{\hat{p}_{i0}/(n_{.j}IJ)\}}} \right]^2.$$

This is expressed as an average over the cells of a weighted distance between the fitted values under the association ($\hat{p}_{ij} = n_{ij}/n_{.j}$) and no-association ($\hat{p}_{i0} = n_{i.}/n_{..}$) models. Uncertainty in the model comparison can then be represented by normal density strips with means \hat{p}_{0j} and standard deviations $\sqrt{\{\hat{p}_{i0}/(n_{.j}IJ)\}}$. The small but clear separation between the two models is apparent and the nature of the effect is clear in the elevated proportion of smokers among the cases. The strength of the evidence expressed is consistent with the p -value of 0.027 arising from the χ^2 -test.

3. Regression

Fig. 5(a) illustrates data on the annual giving in pounds per church member in the dioceses of the Church of England in the early 1980s. Three relevant covariates are also recorded for

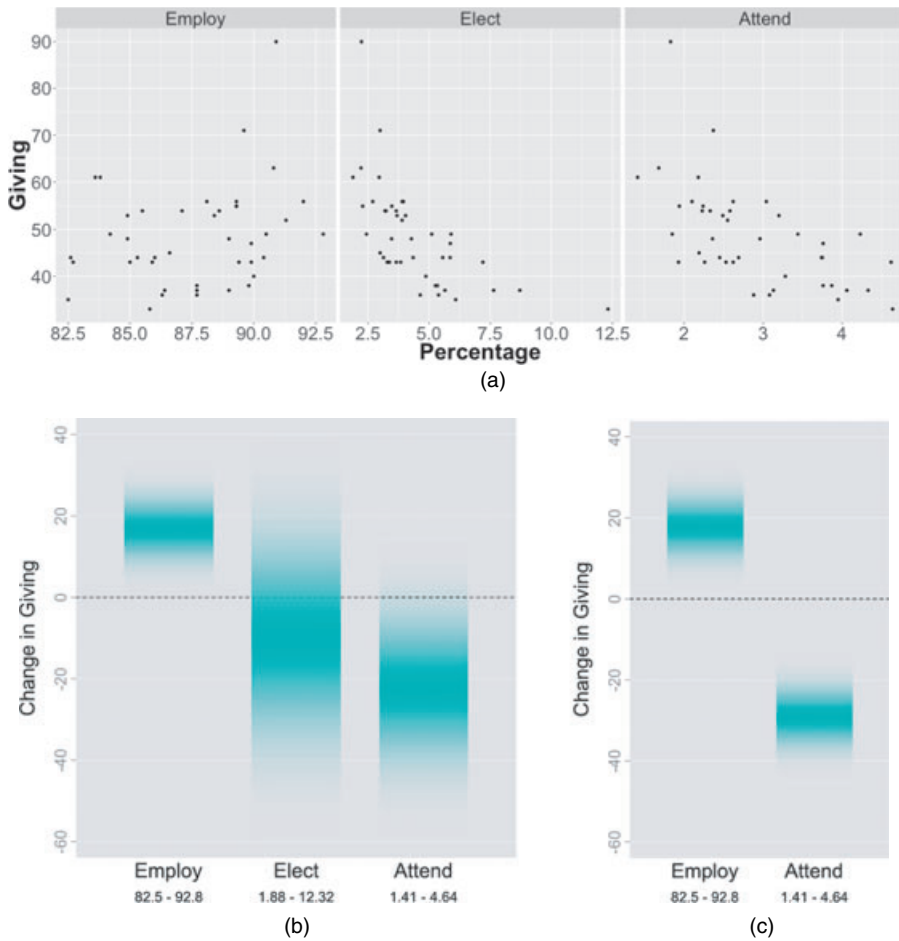


Fig. 5. (a) Data on giving in the Church of England against three covariates and (b), (c) regression effects, scaled to allow direct comparisons, for two models with different selections of covariates, using density strips to indicate the uncertainties involved

each diocese, namely the percentages of the population who are employed (Employ), are on the electoral roll of the church (Elect) and who usually attend church (Attend). Details are available in Pickering (1985).

Regression effects are usually assessed through the sign and size of the regression coefficients. When covariates are measured on very different scales the associated regression parameters are not immediately comparable. A simple device is to plot $\hat{\beta}_i r_i$ for each of the p covariates $i = 1, \dots, p$, where r_i denotes the length of the range of observed values of covariate i . This scaling of the parameter estimates then expresses the change in the response variable across the length of each covariate axis. These values are displayed in Fig. 5(b), using normal density strips centred at $\hat{\beta}_i r_i$ and with standard deviations $\text{se}(\hat{\beta}_i) r_i$ to represent the uncertainty. This enables the regression effects to be directly compared with one another as they are now expressed on the response (Giving) scale in a manner which is naturally associated with the scatter plots, where the primary visual signal lies in the movement of the response across the range of the covariate. Once again, this places the regression effects in the same ‘visual space’ as the original data.

Fig. 5(b) shows a positive association between giving and employment, whereas the associations with membership and attendance are less clear. In fact, these latter two covariates are strongly related to one another, creating an issue with multicollinearity. Fig. 5(c) shows that when only one of these covariates is used, here attendance, then a very strong regression effect is apparent, consistent with the impression that is given by the marginal scatter plots of the observed data. The negative association between giving and attendance is interesting.

4. Curves and surfaces

There are many occasions when the relationship between a response and explanatory variables needs to be modelled in a flexible non-linear manner. Fig. 6 shows an example in the form of a surface estimate constructed from a 'catch score' designed to measure the quantity of marine life found in grabs from the sea bed off the Queensland coast near the Great Barrier Reef. The explanatory variables here are latitude and longitude. The source of the data is referenced in Bowman and Azzalini (1997) where this subset and the construction of suitable surface estimates are described.

One challenging issue is how to display the uncertainty that is associated with this estimate. Bowman (2006) showed how surfaces may be painted with colour to display specific information such as deviation from linearity, but the display of uncertainty in a more general sense is more difficult. Information on standard errors is easily obtainable and so it would be feasible to plot two additional surfaces, defined as 2 standard errors below and above the estimate at every location, but this creates an entirely atypical representation as it employs the extremes of variability at all points simultaneously. It also requires several surfaces to be viewed simultaneously. Some researchers plot a separate surface to represent the standard errors at each location or add further contours for standard errors to the contour plots of the estimated surface. This can be difficult to interpret because the standard errors and the surface estimates are on differ-

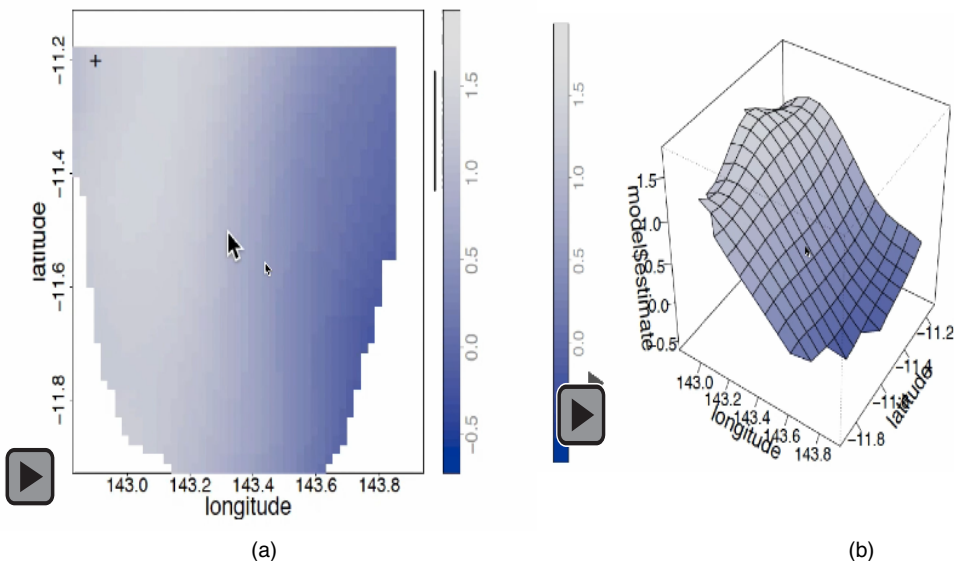


Fig. 6. (a) Contour plot of a surface estimate constructed from the Great Barrier Reef data, with an animation which displays against the colour key the variability intervals attached to specific locations and (b) perspective plot which animates smoothly through surfaces which illustrate the variability of the estimated surface

ent scales or, to use again the expression from Wild *et al.* (2011), they live in different ‘visual spaces’.

The two panels of Fig. 6 propose two solutions to this problem, both using animation. The first is illustrated in Fig. 6(a), which is an interactive plot. As locations are highlighted (in practice by clicking and dragging but here in a prepared animation), a confidence interval, or variability interval as discussed by Bowman and Azzalini (1997), is displayed against the colour key. This enables the user to interrogate uncertainty at any locations of interest and so to build up a picture of the uncertainty pattern across the surface. This strategy does remain in the same visual space but it also retains some of the difficulties in plotting standard errors that were described above and its pointwise nature provides rather partial information.

A more satisfactory approach is to simulate surfaces which conform to the mean and covariance properties of the estimate. This is straightforward to do, as the vast majority of methods of flexible regression have estimates of the form $m = Sy$, where y denotes a vector of response data, S denotes a ‘smoothing matrix’ that is constructed from the values of the covariates and m is a vector of estimated values, which is usually constructed at a regular grid across the surface. Under an assumption of independent errors, an estimate of the error variance σ^2 can also be easily obtained. The details of the estimation process are described by Bowman and Azzalini (1997) in the local linear case and are easily accessible in the literature for other methods. The covariance of the estimated surface points can then easily be estimated as $\Sigma = SS^T \hat{\sigma}^2$, where $\hat{\sigma}^2$ denotes the estimate of error variance. It is then straightforward to simulate surfaces $\{m_i^* : i = 1, \dots\}$ from the multivariate normal distribution $\mathcal{N}(m, \Sigma)$.

The display of a series of unconnected simulated surfaces produces rather abrupt visual transitions and a considerably improved effect is achieved by smoothly tracking between these. It is important to ensure that the intermediate surfaces retain the intended mean and covariance properties. Simple linear interpolation $\alpha m_1^* + (1 - \alpha)m_2^*$, for $0 \leq \alpha \leq 1$, between two simulated surfaces m_1^* and m_2^* does not achieve this as it produces the correct mean m , but an incorrect covariance $\{\alpha^2 + (1 - \alpha)^2\}\Sigma$. A solution is provided by constructing intermediate surfaces as

$$m + \frac{\alpha}{\sqrt{\{\alpha^2 + (1 - \alpha)^2\}}} (m_1^* - m) + \frac{1 - \alpha}{\sqrt{\{\alpha^2 + (1 - \alpha)^2\}}} (m_2^* - m),$$

as these have the correct mean and covariance structure. This is illustrated by the interactive plot in Fig. 6(b). The smooth sequence of surfaces displayed reflects the uncertainty in the estimate in an attractive visual manner. Intuitively, features of the surface which are retained across these simulations may be regarded as systematic rather than the product of sampling variation. For example, there is strong evidence that the plateau nature of the surface at low values of longitude is a real feature. The animation could also have been presented in contour form but the perspective plot is sometimes more effective, especially when combined with interactive control of the viewing angles.

Where a Bayesian analysis is being conducted and m and Σ define a posterior distribution, the interpretation of the simulated surfaces is clear and straightforward. From a frequentist perspective, the issue of bias immediately causes a difficulty in the strict interpretation of confidence regions which is why the terminology variability region is sometimes used, as proposed by Bowman and Azzalini (1997). The sequence of simulated surfaces can then be viewed as Monte Carlo exploration of the confidence ellipsoid that is defined by the mean and covariance matrix. A further interpretation is available as a parametric bootstrap procedure, where simulations are drawn from a fitted model.

Although surfaces offer a more challenging case for the display of uncertainty, it is worthwhile to consider whether the methods that were discussed above might also be used to good effect

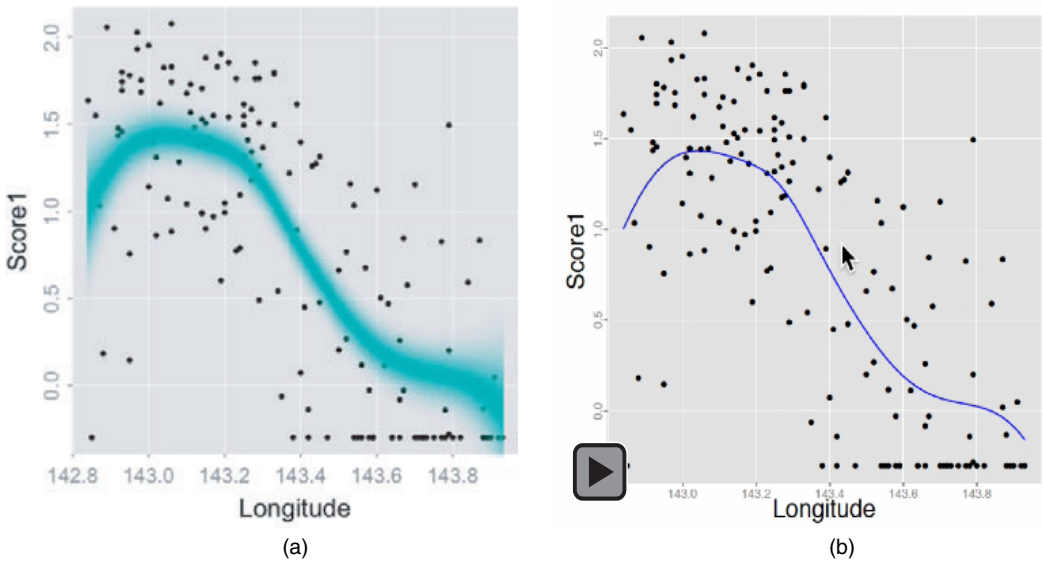


Fig. 7. Plots of catch score against longitude: (a) uncertainty through density strips at a fine grid of values of longitude; (b) animation which displays uncertainty through a smooth sequence of simulated curves

in the context of estimating flexible curves. Fig. 7 displays the data on catch score plotted as a function only of longitude, which is the dominant effect. Fig. 7(a) expresses the uncertainty in the estimation of the relationship between longitude and catch score by displaying a density strip for the position of the curve at a fine grid of values of longitude. This provides a very effective display which avoids the in–out interpretation that is promoted by drawing the end points of confidence bands, as discussed by Jackson (2008) in other examples. The use of a density strip avoids the need to display a single curve estimate at all—a point that was made by Spiegelhalter *et al.* (2011) in other curve estimation settings.

The interactive plot in Fig. 7(b) shows an animation which illustrates the uncertainty through a smooth sequence of simulated curves, produced in exactly the same manner as the surfaces that were discussed above. As in the case of surfaces, the additional information on covariance, which this display incorporates, is very helpful in assessing features of the curve which persist throughout the simulations and which may therefore be regarded as systematic features rather than sampling variation.

Although the examples of this section have involved individual surfaces and curves, the same methods apply to curves and surfaces which are components of more general models, particularly additive models which may involve multiple covariates and model terms.

5. Spatiotemporal data and models

Spatiotemporal data, where measurements of a response of interest are indexed by both space and time, have become very common, leading to considerable research into suitable models. Cressie and Wikle (2011) have provided an excellent introduction to the topic and comment on the lack of suitable graphical methods for exploring spatiotemporal data. The two graphical themes of earlier sections, namely density shading and animation, can also be used to good effect in this setting. Fig. 8 plots data on log-SO₂ pollution across Europe. (A subset of these data was used in Section 2.) Bowman *et al.* (2009) described the data and

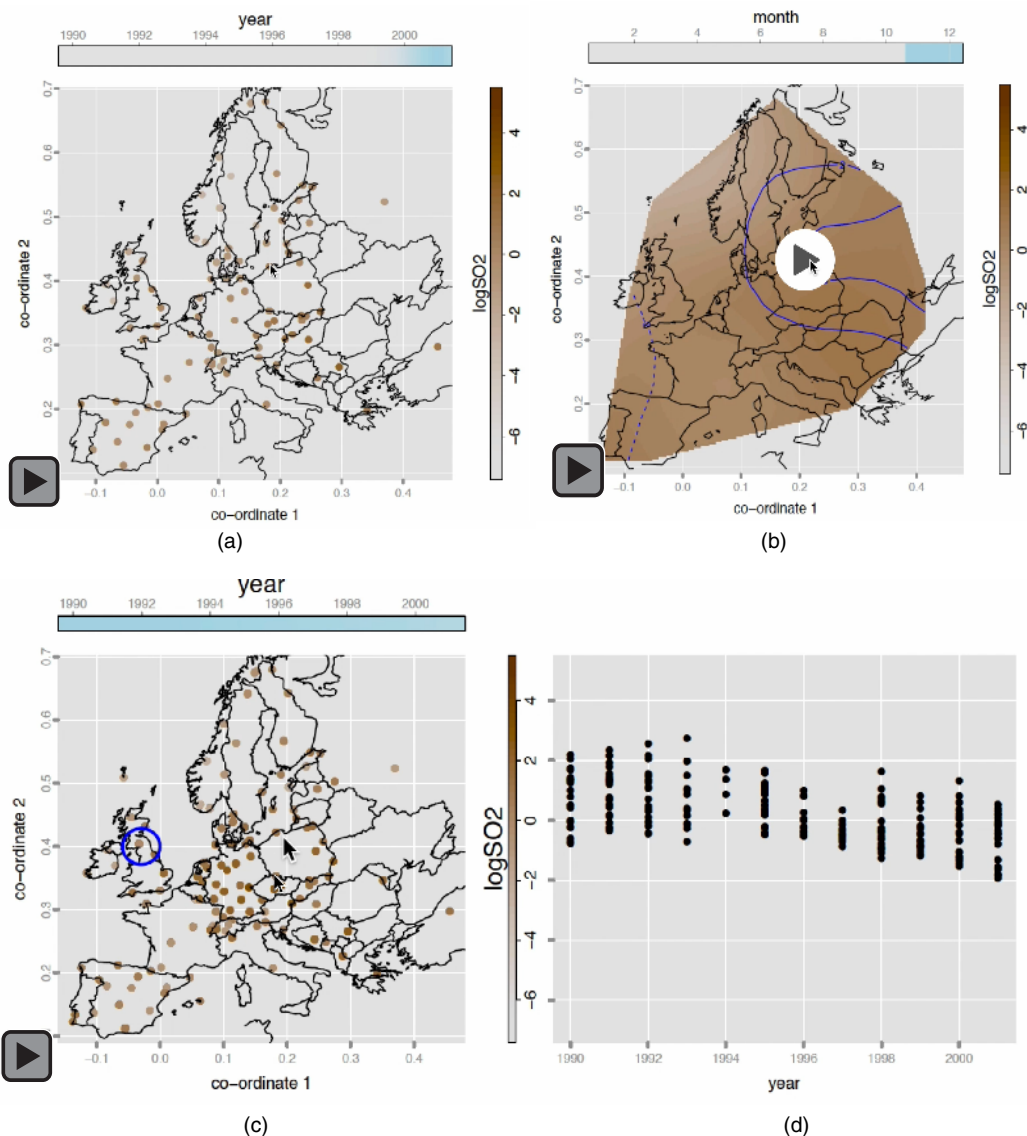


Fig. 8. (a) Animation of SO₂ pollution at spatial locations within specific time windows, (b) animation of the spatial and seasonal terms from a fitted spatiotemporal model and (c) animation of a spatial region (O) for which the pollution values are plotted over time in (d)

constructed a spatiotemporal model involving spatial, temporal and seasonal effects and interactions.

Fig. 8(a) plots the spatial locations of the observations within a specific time window, using colour to indicate the value of the pollution level of each observation. This is an interactive plot which enables the time evolution of the pollution measurements to be explored in a more effective manner than the simultaneous viewing of a set of static plots for selected time windows. The animation employs a time window whose width is indicated in the horizontal bar. The shading that is shown here indicates that the time window is in fact created by a filter, or weight function,

which allows observations to move smoothly into and out of the plotted data. This is achieved by using the hue saturation value form of colour representation (see Manjunath *et al.* (2001)) and downweighting the saturation component according to its distance from the current centre of the time window. The effect of these operations is to create a smooth transition as observations enter and leave the plotted data.

Figs 8(c) and 8(d) show how the time patterns at specific spatial locations can also be explored, where a click on Fig. 8(c) identifies a spatial region within which the pollution values are plotted over time and which may then be dragged across the plotting area. This is a form of interaction with plots known as ‘brushing’ (Becker and Cleveland, 1987) which has been adapted here to the spatiotemporal setting.

Bowman *et al.* (2009) proposed a flexible regression model for log-SO₂, y , with terms involving spatial location, s (two dimensional), time in years, t , and month, z , the last to reflect the seasonal signal. In standard model notation, this can be expressed as

$$y = \mu + m_s(s) + m_t(t) + m_z(z) + m_s(s) : m_t(t) + m_s(s) : m_z(z) + m_t(t) : m_z(z) + \varepsilon,$$

where m denotes a smooth function, ‘:’ denotes interaction terms and ε is an error term. This model was fitted by Bowman *et al.* (2009) through local linear regression and the backfitting algorithm. Here a p -spline representation of each smooth function is used, as described by Eilers and Marx (1996), with 6 and 12 degrees of freedom for one- and two-dimensional terms respectively. The behaviour of the error term ε is modelled by a separable combination of a spherical covariance function $\exp\{-(d_s/\nu)^2\}$ of spatial distance d_s and temporal correlation of auto-regressive AR(1) form on a monthly scale, with correlation parameter ρ . For convenience, the estimated values of $\hat{\nu} = 0.098$ and $\hat{\rho} = 0.569$ that were reported by Bowman *et al.* (2009) are used. After estimation of model terms by penalized likelihood based on independent errors, with estimated standard deviation 0.793, an estimated covariance matrix can then be used to construct adjusted standard errors. Bowman *et al.* (2009) give the details.

Fig. 8(b) shows the interaction of the spatial and seasonal terms $m_s(s) : m_z(z)$. These are the adjustments to an additive model that are required to describe the SO₂ patterns effectively. (This is a case where controls to display the patterns at particular positions are very helpful.) To highlight the need for these adjustments, contours corresponding to 2 or more standard errors from 0 draw attention to the areas where the evidence for interaction is strong. The animation goes on to display the main effects and interaction together: $\mu + m_s(s) + m_z(z) + m_s(s) : m_z(z)$. Here the plot is dominated by the main effects but the contours remain to highlight the presence of the interaction term. This is an example of graphical display involving not only data but also a sophisticated model which can provide clear insight into a complex environmental process.

Fig. 8 was created through the `rp.spacetime` function in the `rpanel` package (Bowman *et al.*, 2007) for R (R Development Core Team, 2013). Jones *et al.* (2014) described software which creates spatiotemporal animations in a convenient automatic manner, specifically designed for the context of groundwater monitoring.

6. Discussion

The graphics that are discussed in the paper aim to provide displays of uncertainty which are intuitive, particularly for a non-technical audience, but which are aligned as closely as possible with the technical construction of the underlying inferential methods. One underlying theme has been the use of colour intensity shading to provide graphics which are more consistent with the fuzzy nature of uncertainty and which counteract the ‘inside–outside’ interpretation of confidence intervals, building on the work of Jackson (2008). A second theme has been the use

of animation which, in particular, enables graphics to remain in the same visual space as the data and model of interest.

Colour selection is an important general issue as this has major implications for the perception of changes across categories or along continuous scales. This is a broad topic which was very helpfully discussed by Zeileis *et al.* (2009).

Acknowledgements

The paper has benefitted greatly from the comments of the reviewers and from conversations with many other colleagues. The research was supported in part by the Economic and Social Research Council through an award (grant ES/L011921/1) to the Urban Big Data Centre at the University of Glasgow.

References

- Becker, R. A. and Cleveland, W. S. (1987) Brushing scatterplots. *Technometrics*, **29**, 127–142.
- Bock, M. T. and Bowman, A. W. (2006) On the measurement and analysis of asymmetry with applications to facial modelling. *Appl. Statist.*, **55**, 77–91.
- Bowman, A. W. (2006) Comparing nonparametric surfaces. *Statist. Modelling*, **6**, 279–299.
- Bowman, A. W. and Azzalini, A. (1997) *Applied Smooth Techniques for Data Analysis*. Oxford: Oxford University Press.
- Bowman, A. W., Crawford, E., Alexander, G. and Bowman, R. W. (2007) rpanel: simple interactive controls for R functions using the tcltk package. *J. Statist. Softw.*, **17**, 1–18.
- Bowman, A. W., Giannitrapani, M. and Scott, E. M. (2009) Spatiotemporal smoothing and sulphur dioxide trends over Europe. *Appl. Statist.*, **58**, 737–752.
- Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2016) Shiny: web application framework for R. *R Package Version 0.13.1*.
- Chen, C.-H., Härdle, W. K. and Unwin, A. (2008) *Handbook of Data Visualization*. Berlin: Springer.
- Cook, D. and Swayne, D. F. (2007) *Interactive and Dynamic Graphics for Data Analysis: with R and GGobi*. New York: Springer.
- Cressie, N. and Wikle, C. K. (2011) *Statistics for Spatio-temporal Data*. New York: Wiley.
- Doll, R. and Hill, A. B. (1950) Smoking and carcinoma of the lung. *Br. Med. J.*, no. 2, 739–748.
- Eilers, P. H. and Marx, B. D. (1996) Flexible smoothing with b-splines and penalties. *Statist. Sci.*, **11**, 89–102.
- Friendly, M. (2000) *Visualizing Categorical Data*. Cary: SAS Institute.
- Gelman, A., Pasarica, C. and Dohdha, R. (2002) Let's practice what we preach: turning tables into graphs. *Am. Statist.*, **56**, 121–130.
- Hartigan, J. A. and Kleiner, B. (1981) Mosaics for contingency tables. In *Computer Science and Statistics: Proc. 13th Symp. Interface* (ed. W. F. Eddy), pp. 268–273. New York: Springer.
- Hood, C., Hosey, M., Bock, M., White, J., Ray, A. and Ayoub, A. (2004) Facial characterization of infants with cleft lip and palate using a three-dimensional capture technique. *Clift Pal. Cranfac J.*, **41**, 27–35.
- Hullman, J., Resnick, P. and Adar, E. (2015) Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLOS One*, **10**, no.11, article e0142444.
- Jackson, C. H. (2008) Displaying uncertainty with shading. *Am. Statist.*, **62**, 340–347.
- Jones, W. R., Spence, M. J., Bowman, A. W., Evers, L. and Molinari, D. A. (2014) A software tool for the spatiotemporal analysis and reporting of groundwater monitoring data. *Environ. Modelling Softw.*, **55**, 242–249.
- Manjunath, B. S., Ohm, J.-R., Vasudevan, V. V. and Yamada, A. (2001) Color and texture descriptors. *IEEE Trans. Circuits Syst. Video Technol.*, **11**, 703–715.
- Pickering, J. (1985) Giving in the Church of England: an econometric analysis. *Appl. Econ.*, **17**, 619–632.
- R Development Core Team (2013) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Sarkar, D. (2008) *Lattice: Multivariate Data Visualization with R*. New York: Springer.
- Seber, G. (1977) *Linear Regression Analysis*. New York: Wiley.
- Spiegelhalter, D., Pearson, M. and Short, I. (2011) Visualizing uncertainty about the future. *Science*, **333**, 1393–1400.
- Theus, M. and Urbanek, S. (2008) *Interactive Graphics for Data Analysis: Principles and Examples*. Boca Raton: Chapman and Hall–CRC.
- Tierney, L. (1988) *XLISP-STAT: a Statistical Environment based on the XLISP Language (Version 2.0)*. Minneapolis: University of Minnesota School of Statistics.
- Tukey, J. W. (1977) *Exploratory Data Analysis*. Reading: Addison-Wesley.

- Tversky, B., Morrison, J. B. and Betancourt, M. (2002) Animation: can it facilitate? *Int. J. Hum. Comput. Stud.*, **57**, 247–262.
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wild, C. J., Pfannkuch, M., Regan, M. and Horton, N. J. (2011) Towards more accessible conceptions of statistical inference (with discussion). *J. R. Statist. Soc. A*, **174**, 247–295.
- Wilkinson, L. (2005) *The Grammar of Graphics*. New York: Springer.
- Zeileis, A., Hornik, K. and Murrell, P. (2009) Escaping rgbland: selecting colors for statistical graphics. *Computnl Statist. Data Anal.*, **53**, 3259–3270.